Contaminants in Aquatic and Terrestrial Environments

TRACING URBAN WASTEWATER CONTAMINANTS INTO THE ATLANTIC OCEAN BY NONTARGET SCREENING

Pablo Antonio Lara-Martín, Aurea C. Chiaia-Hernández, Miriam Biel-Maeso, Rosa María Baena-Nogueras, and Juliane Hollender

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/acs.est.9b06114 • Publication Date (Web): 02 Mar 2020

Just Accepted

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

TRACING URBAN WASTEWATER CONTAMINANTS INTO THE ATLANTIC OCEAN BY NONTARGET SCREENING

Pablo A. Lara-Martín[1*], Aurea C. Chiaia-Hernández[2,3], Miriam Biel-Maeso[1], Rosa M. Baena-Nogueras[1], Juliane Hollender[2,4]

[1]Physical Chemistry Department, Faculty of Marine and Environmental Sciences, University of Cadiz, Campus of International Excellence of the Sea (CEI·MAR), 11510 Puerto Real, Spain

[2]Eawag, Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600 Dübendorf, Switzerland

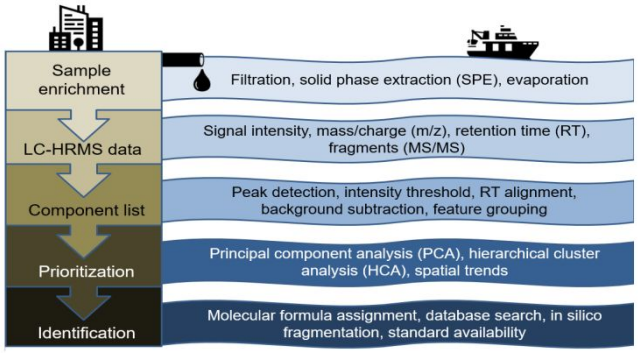
[3]Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, 3012 Switzerland

[4]Institute of Biogeochemistry and Pollutant Dynamics, IBP, ETH Zurich, 8092 Zurich, Switzerland

Abstract

Oceans are the ultimate sink for many of the over 100 million man-made substances. Until now, monitoring was limited to a reduced number of targeted persistent organic pollutants, reaching open waters mainly via atmospheric deposition. However, the composition and fate of the thousands of pollutants reaching the marine environment though wastewater discharges from coastal sources remain largely unexplored. By combining a newly developed non-target screening (NTS) workflow and high resolution mass spectrometry (HRMS), we have identified over 500 sewage-derived contaminants occurring in the ocean. Samples from the NE Atlantic contained this anthropogenic imprint at distances over 50 km from the coastline and > 500 m depth, beyond the continental margin. The range of identified compounds spans from pharmaceuticals and personal care products to food additives and industrial chemicals, including several that have never been reported in the environment, as they escaped conventional targeted analytical methods. Predicting the effects of the continuous input of this chemical "cocktail" on marine ecosystems is a formidable challenge, since 40% of the detected compounds lack information regarding their use and ecotoxicity.

TOC



1. Introduction

Oceans are the vastest ecosystems on Earth, as well as sources of a wide number of resources such as fisheries, minerals, and energy. Coastal zones are the interface between land and the ocean, and have historically attracted human settlements because of these resources, as well as for logistical purposes (e.g., marine trade and transport). The development and utilization of coastal zones have greatly increased during the recent decades, undergoing tremendous socio-economic and environmental changes. The coastal population is projected to be nearly 1 billion by 2030, and most of the world's megacities are located by the coastline (*1*). These rapid changes in the coastal zones have had detrimental impacts towards marine species and habitats. Anthropogenic stressors on marine ecosystems, such as overfishing and nutrient input, are increasing globally (*2*). Among them, pollution by man-made chemicals has recently caught the attention of both environmental scientists and public interest, especially after becoming evident to the naked eye in the form of floating plastic debris in vast areas of the ocean (*3*).

Plastics are not the only artificial compounds mainly derived from land-based sources that can be detected in seawater. As of 2018, there are over 130 million registered chemical substances in the Chemical Abstracts Service (CAS) Registry File. Approximately 400,000 of them are somehow regulated in the international markets (e.g., U.S. Toxic Substances Control Act, or TSCA) and much less regularly monitored in aquatic systems (*4*). The vast number of chemical contaminants that consequently may be potentially present, but have not been detected, in the marine environment poses a great threat to the marine diversity and a challenge for both regulators and researchers. For regulators, one of the most common approaches has been to develop lists of substances to focus on their potential hazard to the ecosystems and/or to human health. In this context, there are several national and international legislations such as the US Clean Water Act (CWA), the European Water Framework Directive (WFD, 2000/60/EC1) and the Marine Strategy Framework Directive (MSFD,

2008/56/EC2) that established a limited number of substances of wide concern (named priority substances) to be regularly monitored in the coastline. Most of the target substances included in these legislations are either metals (As, Cd, Hg, etc.) or persistent organic pollutants (POPs) such as organochlorine compounds, pesticides, brominated flame retardants, perfluorinated chemicals, and recognized endocrine disruptors (e.g., nonylphenol and phthalates). The most comprehensive list of potential chemical contaminants in the marine environment, based on a compilation of substances from relevant aforementioned legislations, global conventions (e.g., Stockholm Convention), and other international organizations (e.g., US EPA) contains approximately 2700 substances (*5*). Many of these proposed compounds are considered as contaminants of emerging concern and have not been implemented yet in routine monitoring programs.

From an analytical point of view, the challenge is even bigger due to the wide range of chemical structures and properties considered and the very low concentrations (< parts per billion, or ppb) occurring in the ocean. To date, most of the effort has been focused on a targeted analysis of selected POPs, susceptible to long-range atmospheric transport and bioaccumulation in marine organisms (*6*). Analysis of non-regulated substances such as specific sewage-markers (e.g., coprostanol) and, most recently, selected pharmaceuticals has often been limited to enclosed seas such as the Baltic Sea and the Mediterranean Sea (*7, 8*). The identification of new contaminants is not only a challenge but also one of the most commonly-mentioned bottlenecks in the advance of disciplines such as environmental chemistry and ecotoxicology (*9*). Recent advances in high resolution mass spectrometry (HRMS), capable of discerning individual chemicals in complex matrices by accurate mass measurement, coupled with chromatographic separation are seen as key tools to address this challenge. By enabling higher throughput analysis of unknown substances at trace levels, the new so-called non-target screening (NTS) strategies aim to capture the universe of anthropogenic organic chemicals occurring in the environment. As we aim to explored the whole chemical composition of samples instead of specific target compounds, complexity in the interpretation of the results increases exponentially. The

resulting big data are managed by crafting analytical workflows capable of analyzing HRMS data with openly accessible online chemical compound databases, structure prediction algorithms, and statistical tools (*10*). To date, much effort has been dedicated to the identification of relevant contaminants and transformation products in wastewater treatment plants (WWTPs) (*11, 12*) and to characterize components in hydraulic fluids used for fracking (*13*).

Here, by using state-of-the-art NTS tools, we present a comprehensive characterization of the oceanic contamination by organic pollutants. As the release of synthetic chemicals into the marine environment may occur through a variety of pathways, including river and atmospheric transport, as well as directly at sea via aquaculture, shipping and fishing activities (*5*), we have focused our attention on sewage-derived contaminants by performing a directed NTS on these substances. Wastewater is the largest contributing source of emerging contaminants to aquatic environments (*14*) and a rising issue in coastal areas due to the aforementioned increase of urban settlements and population. The study area selected was the Gulf of Cadiz (NE Atlantic), as well as its coastline (Cadiz Bay, SW Spain) and the main WWTP (Fig. 1). Comparative analyses of wastewater, coastal waters, and offshore waters collected were performed to unravel the occurrence of 537 wastewater contaminants that could be detected from source to the open ocean, at distances further than 50 km from the coast and > 500 m depth.

2. Materials and Methods

Water sampling. All samples were collected in summer 2015 from the inlet and outlet of the WWTP serving Jerez de la Frontera, the main population in the area, and along 6 different transects: 3 within the Cadiz Bay area (Guadalete River estuary: G1-G10, Sancti Petri creek: S1-9, Rio San Pedro creek: R1-8) and 3 offshore (Cadiz Bay coast: CC1-4, Guadalquivir River mouth: GD1-4. and Trafalgar

Cape: TF1-5) (Fig. 1). Description of the sampling area and further details on how the sampling was conducted are in ref. *15* and Supporting Information (SI), respectively.

Sample extraction and mass spectrometry determination. Briefly, all water samples were filtered with glass fiber filters and extracted using Oasis HLB cartridges. Extracts were spiked with a mix of internal standards (Table S1) and analyzed by high performance liquid chromatography coupled to a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer equipped with an electrospray ionization source (HPLC-ESI-HRMS) and operating in full-scan mode at a mass resolution of 140,000 FHWM referenced to *m/z* 400 according to ref. *16*. For structural elucidation of prioritized compounds (see next section), targeted MS/MS fragmentation was performed at 17,5000 FHWM. Specific information on SPE conditions and HPLC-HRMS instrumental settings can be found in SI.

Selection, prioritization, and identification sewage-derived compounds. A flowchart of the workflow used and the different nodes involved is presented in Fig. 2. The workflow consisted of a combination of nodes implemented with statistical software R (version 3.4.3) and the vendor software Compound Discoverer 2.1. The resulting dataset was subjected to statistical analyses using principal component analysis (PCA) and agglomerative hierarchical cluster analysis (HCA) to prioritize compounds of interest that were present both in wastewater and surface waters samples (Table S2). Identification of the molecular structure of these substances was carried out through a combination of different nodes (Fig. 2), which were complementary and provided identification confidence levels (*17, 18*) between 1 and 4. This process involved using reference standards (EAWAG standards, Table S1), library spectrum data search (mzCloud and MassBank), in-silico fragmentation (MetFrag and FISh), online compound databases (ChemSpider and PubChem) and suspect lists (NORMAN network and Environment and Food Safety, or EFS, lists) (Table S3). All the relevant information regarding the identification of sewage-derived compounds found in both coastal and oceanic waters

(n = 537) is provided in Table S4. Further details on the different workflow nodes, their optimization, and how the statistical analysis was performed can be found in SI.

3. Results and Discussion

Prioritization of sewage-derived compounds. HRMS analysis of wastewater samples at the influent and effluent from the WWTP serving Jerez, the biggest coastal city (215,000 inhabitants) in the study area (Fig. 1), revealed over 250,000 different compounds occurring in sewage. By compounds, we mean the integration of all the different MS features such as molecular ions, salt adducts, and isotopes. They cover a chemical space containing dissolved polar to semi-polar substances within a mass-to-charge (*m/z*) window of 100 to 800 Da that are susceptible to electrospray ionization. Key aspects of the methodology developed here for such integration are provided in SI. To isolate and prioritize signals of interest (sewage-derived contaminants) from the background (naturally occurring components making up the dissolved organic matter), the NTS workflow was first calibrated by using a specific subset of samples. More specifically, we conducted a comparative analysis between the aforementioned wastewater samples and sewage-impacted surface waters from Guadalete Estuary (G1-10) (Fig. 1) to screen for common features. This estuary, highly polluted due to continuous sewage inputs from Jerez WWTP and other pollution sources, has been the subject of a series of studies on targeted contaminants, mainly surfactants (active ingredients in detergent formulations) (*15, 19, 20*).

We found an average number of approximately 47,000 compounds in the samples analyzed along Guadalete Estuary (G1-10). Principal component analysis (PCA) (Fig. S1A in the Supporting Information Appendix) revealed that locations adjacent to the discharge outlet of the WWTP (G1 and G2) differed substantially from the rest. They featured the highest numbers of compounds in the estuary (> 50,000), and most of these compounds had the highest signal intensities. Therefore, we

used the samples from these locations, together with the influent and effluent wastewater samples, to build a preliminary filter for prioritization of sewage-derived substances. The four samples (G1, G2, effluent, and influent) shared 11,256 common compounds. Many of these substances might be organic contaminants in wastewater persistent enough to partially escape sewage treatment and, consequently, be detected in the receiving surface waters. Although the full characterization of the contaminants entering the WWTP is out of the scope of this study, known substances frequently detected in urban wastewater such as caffeine and cocaine (*21*) were identified using certified standards (Table S1). For a considerable fraction of the detected chemicals (35%), relative signal intensities (previously normalized by considering sample volume and ion suppression using internal standards) were higher in the wastewater effluent than in the influent, suggesting either de-conjugation of parent compounds or transformation into different products at the WWTP. Examples are ranitidine S-oxide and dihydroxycarbamazepine, two pharmaceutical biodegradation products also confirmed with available standards.

In a second step, the number of possible sewage-derived contaminants was reduced to 6892 by hierarchical cluster analysis (HCA) of the 11,256 pre-selected compounds using all the samples taken along Guadalete Estuary (Fig. 3A). The main criterion was to select those compounds whose concentrations decreased continuously from G1 (adjacent to the WWTP outlet) towards G10 (at the mouth of the estuary), an indication that they are anthropogenically sourced through wastewater discharges and undergo dilution and/or degradation/sorption during their transport (Fig. 3B). The rest of the substances that did not follow this trend were assigned to either natural dissolved organic matter (34%) (Fig. 3C) or different contamination sources (5%) (Fig. 3D). More specifically, the latter group showed a maximum around sample site G8, where a port serving El Puerto de Santa Maria (90,000 inhabitants) is located, so its occurrence may be associated with the maritime traffic and naval activities occurring in the estuary (*15*). The resulting lists of prioritized sewage-derived components (Table S2) were then included into the NTS workflow to limit the number of substances to detect and

identify in the rest of samples. These include coastal waters from Cadiz Bay (R1-8 and P1-9, at two tidal creeks) and surface and bottom samples taken offshore at the Gulf of Cadiz (CC1-4, GD1-4, and TF1-5, three transects of approximately 50 km length away from the coastline, covering the continental shelf and the shelf break) (Fig. 1).

Identification of wastewater contaminants in the ocean. The average number of possible organic compounds detected in the coastal (n = 26) and oceanic samples (n = 27) analyzed was approximately 36,500 and 12,000, respectively (Fig. 4). The percentage of those compounds included in the list of sewage-derived contaminants was between 1 and 11%, depending on the sample. Specific examples are plotted in Fig. S2, where red dots are wastewater contaminants and the rest (grey dots) are either contaminants from other sources or naturally occurring organic matter. There were two coastal sampling sites, Guadalete Estuary (G1-10) and Rio San Pedro tidal creek (R1-R8) where more than 50,000 substances were measured, indicative of higher concentrations of dissolved organic matter. As discussed in the previous section, this number is partly justified by the continuous input of wastewater from Jerez WWTP into the estuary, which resulted in 6% of total number of compounds, on average, being attributed to sewage-derived chemicals. The tidal creek, however, is not subjected to direct wastewater discharges, as confirmed by the considerably lower percentage (< 2%), number (450-1400), and signal intensities of sewage-derived chemicals detected here. PCA plots (Fig. S1A) also grouped samples from this area differently from the rest. The major contributors to the dissolved organic carbon (DOC) are attributed to effluents from fish farms (*22*), located in the vicinity of sampling site R1. The characterization of the contaminants derived from aquaculture activities is out of the scope of this work.

The spatial distribution of sewage-derived compounds showed a general trend, where the numbers and concentrations of contaminants (expressed as normalized signal intensities by considering sample volume and ion suppression using internal standards) progressively declined when moving from the

main source of wastewater discharge towards the open ocean (Fig. 4). This was expected due to dilution and co-occurrence of different environmental processes such as biodegradation, photolysis or sorption, aimed to reduce the concentrations of chemicals in aquatic environments (*23*). The decreasing trend was not only visible when comparing different coastal and ocean transects, but also to a lesser degree, within each transect as the distance increased from the coastline (Fig. S3). From the almost 7000 compounds included in this analysis (Table S2), the number of sewage-derived substances detected offshore was reduced to between 220 and 600. Some sampling sites (e.g., Trafalgar Cape) were considerably more impacted than the rest (Fig. S1B) possibly due to direct discharges of untreated wastewater from popular touristic spots (e.g., Caños de Meca) not connected to the main sewage network, in agreement to local maximum values for DOC measured there (*24*). Additionally, there were also differences in the vertical distribution of sewage components along the water column. Higher relative signal intensities and numbers were always detected at the surface (Fig. S3), on average 22.8 ± 1.5% higher. These differences have been previously reported for target contaminants such as pharmaceuticals and surfactants in urbanized estuaries (*25*) and can be attributed to positive buoyancy of less dense wastewater plumes entering the ocean via estuaries and marine outfalls.

As the main goal of this study was to characterize substances entering and occurring in the Atlantic Ocean, identification of compounds by assigning specific compound formulas, structures, and names was performed uniquely for those detected in offshore samples. This resulted in 537 different sewage-related chemicals (Table S4). Confidence identification levels from 1 (the highest confidence, confirmed with reference standards) to 4 (the lowest, tentative chemical compound from a chemical database) were assigned by using a combination of available chemical pure standards, mass spectra libraries, online chemical database searches, and in-silico mass spectral fragmentation algorithms (see Materials and Methods section). Some examples to illustrate different aspects used for identification are shown in Fig. S4. Here, the artificial sweetener sucralose, the food additive sulfurol, and the

industrial chemical 17-amino-3,6,9,12,15-pentaoxaheptadecan-1-ol, were identified at levels 1, 2, and 3, respectively. Identification confidence levels and elemental composition of all the sewage-derived contaminants detected in oceanic waters from the Gulf of Cadiz is summarized in Fig. 5A, where information on their inclusion in chemical databases and known applications is presented in Fig. 5B. The vast majority of contaminants (85%) were exclusively formed of C, H, N, O, and S atoms, covering a wide range of functional groups including alcohols, carboxylates, alkylamines, sulfonated and sulfated compounds. The rest were either halogenated chemicals (mainly chlorinated substances, see blue bar in Fig. 5A) or had P and/or Si atoms in their composition (see red bar in Fig. 5A). Other chemicals such as hydrocarbons (composed of C and H only) were not detected due to the limitations of the analytical approach used, based on HPLC-ESI-HRMS.

Only about 20% of the prioritized sewage contaminants could be identified with high degree of confidence (levels 1 and 2, see light green and blue sectors, respectively, in Fig. 5A) due to the limited number of standards available in our laboratory (Table S1) and the still relatively low (although growing) number of HRMS spectra recorded from commercially available standards. For instance, as of 2018, there were approximately 16 thousand substances in one of the mass spectral libraries used here (MassBank); this number represents a tiny fraction of the more than 130 million registered chemical substances in the Chemical Abstracts Service (CAS). Our identification effort still resulted in over 100 chemicals positively identified in the ocean, many of them never reported before. The bulk of the prioritized list (60% of the substances), however, was composed by chemicals identified at level 3 by combining database searches (suspect lists in Table S3, ChemSpider, and PubChem) and in-silico fragmentation tools (MetFrag and FISh). Higher scores in Table S4 (columns P, R, and S) are indicative of higher probabilities of achieving successful identification. For the remaining chemicals (17%), HRMS information was not enough to elucidate their chemical structures. Tentative names were assigned from top-ranked compounds in ChemSpider and PubChem databases (in terms of number of references) that matched their molecular formula.

Environmental relevance of the chemicals identified. We could find 95% of the different substances identified in the ocean in online databases (ChemSpider and PubChem, light blue and light green sectors combined in Fig. 5B), and almost 50% of the possible candidates were already compiled in lists of suspected environmental contaminants (EFS and NORMAN network lists, see Table S3) (light green sector in Fig. 5B). Relevant environmental information such as their uses and applications was found for over 60% of them. This allowed their classification into six different groups of chemicals: pharmaceuticals and related bioactive substances, pesticides, personal care products, surfactants, industrial chemicals, and food additives. The top-ten compounds of each group, in terms of signal intensity and frequency of detection, are presented in Fig. 6. Their relative abundance is also depicted as maximum signal intensity ranges in coastal (G1-10, R1-8, and P1-9, red bars) and oceanic (CC1-4, GD1-4, and TF1-5, blue bars) samples, the latter often being several orders of magnitude lower. The rest of the contaminants were not considered for the analysis shown in Fig. 6 as no data on their possible uses and/or ecotoxicological relevance could be retrieved (red bar in Fig. 5B).

Pharmaceutically active compounds were, in terms of number of identified chemicals (n = 144), the most relevant group of contaminants, comprising up to 25% of the prioritized wastewater components detected in the NE Atlantic. Those compounds showing highest intensities were analgesic (e.g., antipyrine, also known as phenazone) and psychiatric drugs such as venlafaxine (an antidepressant), carbamazepine, and olanzapine. The former group were reported worldwide at the highest concentrations in coastal WWTPs due to their higher consumption (non-prescription drugs in many cases), whereas the latter are very recalcitrant when undergoing conventional sewage treatment (*25*). The presence of metabolites (4-formylaminoantipyrine, 4-acetamidoantipyrine, and valsartan acid, among others) and lidocaine (a local anesthetic) was also confirmed (level 1), whereas antivirals such as telvibudine and nevirapine were identified at level 3. Possible effects on marine biota derived from the occurrence of these bioactive substances have been recently reviewed by Mezzelani and co-

307 workers (*26*), including alterations in the enzymatic and immunological systems of filter feeders. As
308 NTS workflows are not aimed for accurate quantification of the concentrations, we proceeded to re-
309 analyze our samples using specific methodology for target determination of pharmaceuticals. Total
310 concentrations of these contaminants were between 50 and 200 ng L^{-1}, not currently known to be high
311 enough to pose any foreseen risk towards marine biota (*20*). This preliminary environmental risk
312 assessment, however, was not without limitations due to poor ecotoxicological data regarding the
313 toxicity of these substances towards marine species, as well as unknown additive or synergistic effects
314 with other chemicals co-existing in the water column. Among these chemicals, pesticides are also
315 bioactive substances on their own, designed to kill specific organisms. Fungicides, such as metalaxyl
316 and carbendazim, and herbicides (terbutryn, 2,4-D, and fluometuron) were identified at different
317 confidence levels. Toxicity towards primary producers such as marine microalgae can be expected
318 even at sub-ppb concentrations (*27*).
319
320 Surfactants were second in number of identified sewage-derived components, but first in terms of
321 signal intensities. Although their removal in conventional WWTPs is very efficient (*25*) (> 95%),
322 their current worldwide production is the highest of all synthetic chemicals (after plastics), with over
323 15 million tonnes per year (*28*). As a consequence, these compounds are reported at the highest
324 concentrations in treated effluents and streams in comparison with other organic contaminants not so
325 efficiently eliminated during wastewater treatment but used in lower amounts (e.g., psychiatric drugs)
326 (*29*). Linear alkylbenzene sulfonates (LAS), polyethylene glycols (PEG), and polypropylene glycols
327 (PPG) were the main components detected in seawater, together with their respective biodegradation
328 products (e.g., sulfophenyl carboxylic acids, or SPCs, and oxidized PEG) and manufacturing
329 impurities (dialkyltetralin sulfonates, or DATS). Rather than individual compounds, they are complex
330 mixtures comprising different homologous and ethoxymer series (Fig. S5) with changing
331 environmental and ecotoxicological properties (*30*). Although readily biodegradable according to
332 screening tests, their elimination was shown to be considerably slower under marine conditions (*31*)

in comparison with freshwater environments. This is explained by a combination of marine microbial communities being less active than their freshwater counterparts toward xenobiotic chemicals and complexation with ions in seawater that reduces bioavailability, thereby inhibiting biodegradation (*30*). Surfactants are used in a myriad of applications including cleaning detergents (LAS), personal care products (PEG and PPG), and various industrial uses (e.g., oil, textiles, polymers, agriculture, and paints). The global surfactant market is forecast to grow at a compound annual rate of 6% over the next years, with rising demand for personal care products (PCPs) being the market driver (*30*). Other chemicals included in the formulation of PCPs that were detected in our samples were fragrances (galaxolidone, a degradation product from galaxolide), UV filters (sulisobenzone), and insect repellents (DEET and picaridin). Some of these compounds, reported in freshwater resources from all continents, are bioaccumulative and suspected endocrine disruptors, leading to potential alterations on fecundity, growth and development of exposed species (*14*).

Lastly, industrial chemicals were the third most relevant group identified by our NTS workflow. Their inclusion in the prioritized list of sewage-derived contaminants can be explained by the presence of industrial activities (e.g., wineries, laundry services, food processing, auto repair shops, etc.) within the city and its surroundings, also served by Jerez WWTP. This is a very diverse group comprising substances used in a wide range of applications, including processing aids (siloxanes), buffering agents (CAPSO), reaction intermediates (benzenesulfonamide), complexing agents (21-Crown-7), and corrosion inhibitors (benzotriazoles). The vast majority of the chemicals included in this group (n = 54) has never been reported in the marine environment and there is no literature on their risks towards marine species. This includes food additives such as the flavoring agents sulfurol, identified at level 2, geranyl acetate, and eugenol benzoate (level 3). Some of them are not only synthetically produced, but also natural products. Such is the case of panthenol, a vitamin (B5) but also a moisturizer widely used in PCPs that was recently identified in WWTP using a NTS approach (*12*) and also detected in our seawater samples (Table S4). To discern between natural and anthropogenic

359 sources and the relative contributions of these sources is challenging and certainly would not be
360 possible without using additional analytical techniques (e.g., compound-specific stable isotope
361 analysis). Nevertheless, the highest concentrations observed in urban wastewater ensure that cities
362 are one of the main sources of this type of chemicals towards the marine environment. Within this
363 category, only non-volatile cyclosiloxanes (D6 and D7), also used in PCPs and with potential
364 carcinogenicity, and benzotriazoles were previously reported in WWTP effluents (*32, 33*). Both are
365 resistant to biodegradation and very persistent in aquatic environments, and the latter have been
366 proposed as anthropogenic indicators in groundwater and rivers (*34*). Similarly, the artificial
367 sweetener sucralose, the food additive detected at the highest intensity in oceanic samples, is
368 considered as an ideal marker of sewage contamination in freshwater (*35*) and, more recently, coastal
369 systems (*36*) due to its recalcitrance, source specificity, and high solubility. Concentrations up to 70
370 ng L^{-1} were reported in the Gulf Stream (*37*), the only available data reported from offshore samples.
371 In the present study, we determined an average 350-fold decrease in the relative abundance of these
372 well-known sewage markers (Fig. S6A) from the most contaminated station (G1) to offshore samples.
373 Due to the low reactivity of these markers, it was mostly attributed to dilution of wastewater in the
374 ocean. Other chemicals more prone to undergo degradation (e.g., antipyrine and its main metabolites)
375 showed a decrease in their concentration of more than 1000-fold from the estuary to the ocean. The
376 transformation of the parent compound into different degradation products during their transport was
377 evident as the relative abundances increased for the latter as we moved away from the contamination
378 source (Fig. S6B).
379
380 **Future perspectives.** This work presents a list with over 500 different contaminants detected in
381 coastal and oceanic samples from NE Atlantic. Their occurrence not only along the coastline but also
382 in the whole continental shelf and beyond in a very dynamic environment, as opposed to previous
383 targeted studies in enclosed seas, confirms their ubiquity. The detection of these chemicals in the
384 open sea can only be explained by a combination of persistence in seawater and continuous input of

385 significant amounts from land-based sources such as WWTPs. Immediate efforts should be focused
386 towards development of specific target methodologies aimed at the accurate determination of the
387 concentrations of those contaminants detected by NTS, as well as ecotoxicity assays and/or modeling
388 to discern whether they may pose risks towards marine biota. This is also a long-term task as the
389 number of synthetic chemicals is continuously increasing (e.g., in 2015, 56 new drugs were launched
390 and more of 7000 compounds were in trial) (*38*). Their global trends in terms of worldwide production
391 and diversification has already outpaced the rates of change of other agents of global change such as
392 rising CO_2 atmospheric concentrations and habitat destruction (*39*), but pollution by synthetic
393 chemicals has not yet been included in most analyses of global change. In addition, and in spite of
394 nowadays stricter environmental policies, concentrations of many sewage-derived chemicals in the
395 marine environment are expected to keep rising within the next decades as the human population
396 increases while conventional WWTPs are incapable of efficient removal of many contaminants. This
397 growing trend has already been confirmed for pharmaceuticals from recent targeted analyses of dated
398 marine sediment cores (*40*). Mitigation measures include improving wastewater treatment through
399 the use of additional technologies (e.g., advanced oxidation processes) in current and planned coastal
400 WWTPs, which have proved to increase water quality in sewage-impacted freshwater settings (*41*).
401
402 This study also explores the boundaries of the current existing methodologies for environmental high-
403 throughput identification of compounds by HRMS. Although targeted analyses using reference
404 standards remains the most reliable way to determine contaminants, it is not feasible to extend this
405 approach over more than a few hundreds of substances of interest (more than 600 were used here) for
406 routine monitoring. As mass spectral libraries are growing and computational mass spectrometry is
407 advancing at a fast pace, the confidence in the identification of unknown substances using NTS
408 workflows will increase in the near future. Already acquired HRMS data such as ours (available in
409 File S1) will also benefit from upcoming achievements in the field (e.g., more robust in-silico
410 fragmentation algorithms and identification of new contaminants) through the use of retrospective

screening, as has been recently tested with aqueous environmental samples from different European countries (*42*). Use of complementary techniques such as GC-HRMS for screening of hydrophobic and semi-volatile chemicals will allow to expand the analytical horizon towards an even more complete picture of the chemical universe in our environment (*43*). Lastly, by applying NTS workflows similar to the one described here, different potential contamination sources towards the marine environment such as aquaculture and maritime traffic could be investigated to unravel additional contaminants.

Associated content

Supporting Information. Materials and methods, Figures S1-S7, Tables S1-S4, and File S1 (MS/MS spectra). This material is available free of charge via the Internet at http://pubs.acs.org.

Author information

***Corresponding Author.** Address: Campus Río San Pedro s/n, Campus de Excelencia Internacional del Mar (CEI·MAR), 11510 Puerto Real, Cádiz, Spain. Phone: +34 956 016159. E-mail: pablo.lara@uca.es

Notes. The authors declare no competing financial interest.

Acknowledgments

Thanks to J.E Schollé and K. Kiefer for their HPLC-HRMS and MetFragR support, respectively, and to S. Salzmann during the revision process. This research was supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) (CTM2015-70731-R) and by the University of Bern through the Interfaculty Research Cooperation One Health.

436 # References

437

438 1. Neumann, B.; Vafeidis, A.T.; Zimmermann, J.; Nicholls, R.J. Future coastal population growth
439 and exposure to sea-level rise and coastal flooding – A global assessment. *PLoS ONE* **2015**, 10, 3.
440 2. Halpern, B.S.; Frazier, M.; Potapenko, J.; Casey, K.S.; Koenig, K.; Longo, C.; Lowndes, J.S.;
441 Rockwood, R.C.; Selig, E.R.; Selkoe, K.A.; Walbridge, S. Spatial and temporal changes in
442 cumulative human impacts on the world's ocean. *Nat. Commun*. **2015**, 6, 7615.
443 3. Cózar, A.; Echevarría, F.; González-Gordillo, J.I.; Irigoien, X.; Úbeda, B.; Hernández-León, S.;
444 Palma, A.T.; Navarro, S.; García-de-Lomas, J.; Ruiz, A.; Fernández-de-Puelles, M.L.; Duarte, C.M.
445 Plastic debris in the open ocean. *PNAS* **2014**, 111, 10239-10244.
446 4. CAS, Database Counter, 2018, http://www.cas.org/content/counter (accessed 10.10.18).
447 5. Tornero, V; Hanke, G. Potential chemical contaminants in the marine environment: An overview
448 of main contaminant lists. ISBN 978-92-79-77045-6, EUR 28925, **2017**.
449 6. Luek, J.L.; Dickhut, R.M.; Cochran, M.A.; Falconer, R.L.; Kylin, H. Persistent organic pollutants
450 in the Atlantic and southern oceans and oceanic atmosphere. *Sci. Total Environ.* **2017**, 583, 64-71.
451 7. Maldonado, C; Dachs, J.; Bayona, J.M. Trialkylamines and Coprostanol as Tracers of Urban
452 Pollution in Waters from Enclosed Seas: The Mediterranean and Black Sea. *Environ. Sci. Technol.*
453 **1999**, 33, 3290-3296.
454 8. Björlenius, B.; Ripszám, M.; Haglund, P.; Lindberg, R.H.; Tysklind, M.; Fick, J. Pharmaceutical
455 residues are widespread in Baltic Sea coastal and offshore waters - Screening for pharmaceuticals
456 and modelling of environmental concentrations of carbamazepine. *Sci. Total Environ.* **2018**, 633,
457 1496-1509.
458 9. Schwarzenbach, R.P.; Escher, B.I.; Fenner, K.; Hofstetter, T.B.; Johnson, C.A.; von Gunten, U.;
459 Wehrli, B. The challenge of micropollutants in aquatic systems. *Science* **2006**, 313, 1072-1077.

10. Hollender, J.; Schymanski, E.L.; Singer, H.; Ferguson, P.L. Non-target screening with high resolution mass spectrometry in the environment: Ready to go?. *Environ. Sci. Technol.* **2017**, 51, 11505-11512.

11. Schollée, J.E.; Schymanski, E.L.; Avak, S.E.; Loos, M.; Hollender, J. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* **2015**, 87, 12121−12129.

12. Gago-Ferrero, P.; Schymanski, E.L.; Bletsou, A.A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N.S. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* **2015**, 49, 12333-12341.

13. Hoelzer, K.; Summer, A.J.; Karatum, O.; Nelson, R.K.; Drollette, B.D.; O'Connor, M.P.; D'Ambro, E.L.; Getzinger, G.J.; Lee Ferguson, P.; Reddy, C.M.; Elsner, M.; Plata, D.L. Indications of Transformation Products from Hydraulic Fracturing Additives in Shale-Gas Wastewater. *Environ. Sci. Technol.* **2016**, 50, 8036-8048.

14. Montes-Grajales, D.; Fennix-Agudelo, M.; Miranda-Castro, W. Occurrence of personal care products as emerging chemicals of concern in water resources: A review. *Sci. Total Environ.* **2017**, 595, 601-604.

15. Lara-Martín, P.A.; Gómez-Parra, A.; González-Mazo, E. Sources, transport and reactivity of anionic and non-ionic surfactants in several aquatic ecosystems in SW Spain: A comparative study. *Environ. Poll.* **2008**, 156, 36-45.

16. Chiaia-Hernández, A.; Günthardt, B.F.; Frey, M.P.; Hollender, J. Unravelling Contaminants in the Anthropocene Using Statistical Analysis of Liquid Chromatography−High-Resolution Mass Spectrometry Nontarget Screening Data Recorded in Lake Sediments. *Environ. Sci. Technol.* **2017**, 51, 12547-12556.

17. Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, 48, 2097−2098.

18. Ruttkies, C.; Schymanski, E.L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminform.* **2016**, 8, 3.

19. Corada-Fernández, C.; Lara-Martín, P.A.; Candela, L.; González-Mazo, E. Tracking sewage derived contamination in riverine settings by analysis of synthetic surfactants. *J. Environ. Monitor.* **2011**, 13, 2010-2017.

20. Biel-Maeso, M.; Baena-Nogueras, R.M.; Corada-Fernández, C.; Lara-Martín, P.A. Occurrence, distribution and environmental risk of pharmaceutically active compounds (PhACs) in coastal and ocean waters from the Gulf of Cadiz (SW Spain). *Sci. Total Environ.* **2018**, 612, 649-659.

21. Choi, P.M.; Tscharke, B.J.; Donner, E.; O'Brien, J.W.; Grant, S.C.; Kaserzon, S.L.; Mackie, R.; O'Malley, E.; Crosbie, N.D.; Thomas, K.V.; Mueller, J.F. Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC* **2018**, 105, 453-469.

22. Burgos, M.; Ortega, T.; Bohórquez, J.; Corzo, A.; Rabouille, C.; Forja, J.M. Seasonal variation of early diagenesis and greenhouse gas production in coastal sediments of Cadiz Bay: Influence of anthropogenic activities. *Estuar. Coast. Shelf. Sci.* **2018**, 200, 99-115.

23. Jaffé, R. Fate of hydrophobic organic pollutants in the aquatic environment: A review. *Environ. Pollut.* **1991**, 69, 237-257.

24. González-Ortegón, E.; Amaral, V.; Sánchez-Leal, R.F.; Bellanco, M.J.; Jiménez, M.P.; Forja, J.; Vilas, C.; Tovar-Sanchez, A. Sources and coastal distribution of dissolved organic matter in the Gulf of Cadiz. *Sci. Total Environ.* **2018**, 630, 1583-1595.

25. Lara-Martín, P.A.; González-Mazo, E.; Petrovic, M.; Barceló, D.; Brownawell, B.J. Occurrence, distribution and partitioning of nonionic surfactants and pharmaceuticals in the urbanized Long Island Sound Estuary (NY). *Mar. Pollut. Bull.* **2014**, 85, 710-719.

26. Mezzelani, M.; Gorbi, S.; Regoli, F. Pharmaceuticals in the aquatic environments: Evidence of emerged threat and future challenges for marine organisms. *Mar. Environ. Res.* **2018**, 140, 41-60.

510 27. Moro, L.; Pezzotti, G.; Turemis, M.; Sanchis, J.; Farré, M.; Denaro, R.; Giacobbe, M.G.; Crisafi,
511 F.; Giardi, M.T. Fast pesticide pre-screening in marine environment using a green microalgae-based
512 optical bioassay. *Mar. Pollut. Bull.* **2018**, 129, 212-221.
513 28. Acmite Market Intelligence, 2016. Market Report: Global Surfactant Market, fourth ed., **2016**.
514 29. Kolpin, D.W.; Furlong, E.T.; Meyer, M.T.; Thurman, E.M.; Zaugg, S.D.; Barber, L.B.; Buxton,
515 H.T. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams,
516 1999−2000: A national reconnaissance. *Environ. Sci. Technol*. **2002**, 36, 1202−11.
517 30. Jackson, M.; Eadsforth, C.; Schowanek, D.; Delfosse, T.; Riddle, A.; Budgen, N. Comprehensive
518 review of several surfactants in marine environments: Fate and ecotoxicity. *Environ. Toxicol. Chem.*
519 **2016**, 35, 1077-1086.
520 31. Perales, J.A.; Manzano, M.A.; Garrido, M.C.; Sales, D.; Quiroga, J.M. Biodegradation kinetics
521 of linear alkylbenzene sulphonates in sea water. *Biodegradation* **2007**, 18, 63-70.
522 32. Giger, W.; Schaffner, C.; Kohler, H.P. Benzotriazole and tolyltriazole as aquatic contaminants.
523 1. Input and occurrence in rivers and lakes. *Environ. Sci. Technol.* **2006**, 40, 7186-7192.
524 33. Capela, D.; Ratola, N.; Alves, A.; Homem, V. Volatile methylsiloxanes through wastewater
525 treatment plants – A review of levels and implications. *Environ. Int.* **2017**, 102, 9-29.
526 34. Seit, W.; Winzenbacher, R. A survey on trace organic chemicals in a German water protection
527 area and the proposal of relevant indicators for anthropogenic influences. *Environ. Monit. Assess*
528 **2017**, 189, 244.
529 35. Lange, F.T.; Scheurer, M.; Brauch, H.J. Artificial sweeteners--a recently recognized class of
530 emerging environmental contaminants: a review. *Anal. Bioanal. Chem.* **2012**, 403, 2503-2518.
531 36. Baena-Nogueras, R.M.; Traverso-Soto, J.M.; Biel-Maeso, M.; Villar-Navarro, E.; Lara-Martín,
532 P.A. Sources and trends of artificial sweeteners in coastal waters in the bay of Cadiz (NE Atlantic).
533 *Mar. Pollut. Bull.* **2018**, 13, 607-616.

534 37. Mead, R.N.; Morgan, J.B.; Avery, G.B.; Kieber, R.J.; Kirk, A.M.; Skrabal, S.A.; Willey, J.D.
535 Occurrence of the artificial sweetener sucralose in coastal and marine waters of the United States.
536 *Mar. Chem.* **2009**, 116, 13–17.
537 38. International Federation of Pharmaceutical Manufacturers, Associations (IFMPA). The
538 Pharmaceutical Industry and Global Health: Facts and Figures 2017. ISBN 978-2-940498-50-5, **2017**.
539 39. Berhnard, E.S.; Rosi, E.J.; Gessner, M.O. Synthetic chemicals as agents of global change. *Front.*
540 *Ecol. Environ.* **2017**, 15, 84-90.
541 40. Lara-Martín, P.A.; Renfro, A.A.; Cochran, J.K.; Brownawell, B.J. Geochronologies of
542 Pharmaceuticals in a Sewage-Impacted Estuarine Urban Setting (Jamaica Bay, New York). *Environ.*
543 *Sci. Technol*. **2015**, 49, 1202−1211.
544 41. Salimi, M.; Esrafili, A.; Gholami, M.; Jonidi Jafari, A.; Rezaei Kalantary, R.; Farzadkia, M.;
545 Kermani, M.; Sobhi, H.R. Contaminants of emerging concern: a review of new approach in AOP
546 technologies. *Environ. Monit. Assess.* **2017**, 189, 414.
547 42. Alygizakis, N.A.; Samanipour, S.; Hollender, J.; Ibáñez, M.; Kaserzon, S.; Kokkali, V.; van
548 Leerdam, J.A.; Mueller, J.F.; Pijnappels, M.; Reid, M.J.; Schymanski, E.L.; Slobodnik, J.; Thomaidis,
549 N.S.; Thomas, K.V. Exploring the Potential of a Global Emerging Contaminant Early Warning
550 Network through the Use of Retrospective Suspect Screening with High-Resolution Mass
551 Spectrometry. *Environ. Sci. Technol.* **2018**, 52, 5135-5144.
552 43. Hernández, F.; Ibáñez, M.; Portolés, T.; Cervera, M.I.; Sancho, J.V.; López, F.J. Advancing
553 towards universal screening for organic pollutants in waters. J. Hazard. Mater. 2015, 282, 86-95.
554
555
556
557
558
559
560

561 # Figure Captions

562

563 **Fig. 1.** Map showing the location of the Gulf of Cadiz (SW Spain) and sampling sites: GD1-4 =
564 Guadalquivir River mouth transect, CC1-4 = Cadiz Coast transect, and TF1-5 = Trafalgar Cape
565 transect. The inset shows the Cadiz Bay and sampling sites: G1-10 = Guadalete Estuary transect,
566 R1-8 = Rio San Pedro transect, and P1-9 = Sancti Petri transect. The location of the main WWTP is
567 also depicted (Schlitzer, R. Ocean Data View, odv.awi.de, 2020).

568

569 **Fig. 2.** Workflow tree and nodes involved in the non-target screening (NTS) and identification of
570 sewage-derived compounds in coastal and oceanic samples. Maximum confidence identification level
571 provided by different nodes is also depicted: 1 (chemical standards), 2 (spectral libraries), 3 (in-silico
572 fragmentation algorithms), and 4 (online compound databases and suspect lists). A detailed
573 explanation of each node can be found in Supporting Information.

574

575 **Fig. 3.** Hierarchical cluster analysis (HCA) of potential sewage-derived compounds (n = 11,256)
576 along the Guadalete Estuary (G1-10) (A) and selected clusters: potential contaminants from Jerez
577 WWTP (61%) (B), naturally occurring organic matter (34%) (C), and other pollution sources (5%)
578 (D). Examples of compounds for each cluster are plotted in red.

579

580 **Fig. 4.** Box plot showing the variability (minimum, first quartile, median, third quartile, and
581 maximum) in the number of all organic compounds and potential sewage-derived contaminants
582 detected in sampling stations at the Gulf of Cadiz (GD1-4 = Guadalquivir River mouth transect, CC1-
583 4 = Cadiz Coast transect, and TF1-5 = Trafalgar Cape transect) and the Cadiz Bay (P1-9 = Sancti
584 Petri transect, R1-8 = Rio San Pedro transect, and G1-10 = Guadalete Estuary transect). Normalized
585 signal intensity ranges of sewage-derived contaminants are also plotted.

586

587 **Fig. 5.** Characterization of the sewage-derived contaminants (n = 537) detected in oceanic waters
588 from the Gulf of Cadiz: identification confidence levels (1-4, sector graph) and elemental composition
589 (bar graph) (A), and percentage of these chemicals found in online databases and suspect lists (sector
590 graph) and known uses and applications (bar graph) (B). For those compounds included in the red
591 N/A sector and bar in Fig. 5B, information was not found on their identity or uses/applications,
592 respectively.

593

594 **Fig. 6.** Maximum normalized signal intensities of the top 10 sewage-derived contaminants in coastal
595 (Cadiz Bay) and oceanic (Gulf of Cadiz) waters sorted by use/application class. Confidence
596 identification levels (1 = green, 2 = blue, 3 = orange) are presented on the right of the compound
597 names.

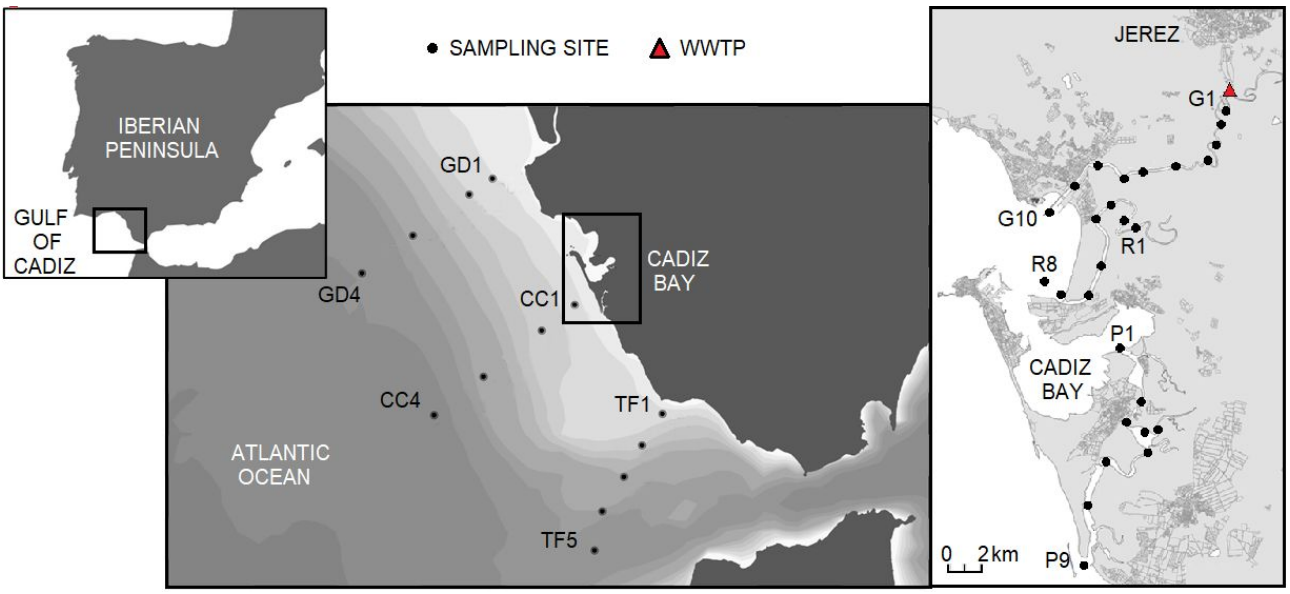

SAMPLING SITE
WWTP
IBERIAN PENINSULA
GULF OF CADIZ
GD1
GD4
CC1
CC4
TF1
TF5
CADIZ BAY
ATLANTIC OCEAN
JEREZ
G1
G10
R1
R8
P1
CADIZ BAY
P9
0 2km

Fig. 1

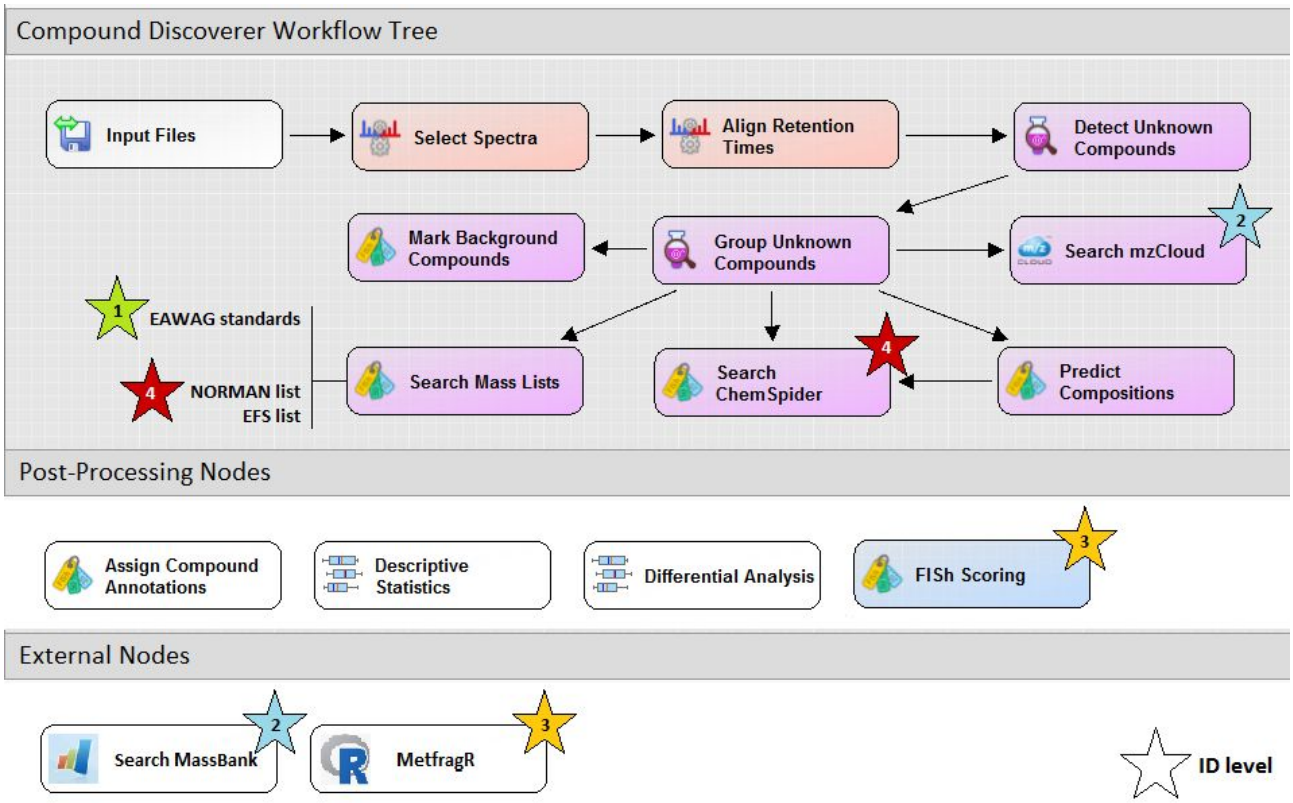

Compound Discoverer Workflow Tree
Input Files
Select Spectra
Align Retention Times
Detect Unknown Compounds
Mark Background Compounds
Group Unknown Compounds
Search mzCloud
2
1
EAWAG standards
4
NORMAN list
EFS list
Search Mass Lists
Search ChemSpider
4
Predict Compositions
Post-Processing Nodes
Assign Compound Annotations
Descriptive Statistics
Differential Analysis
FISh Scoring
3
External Nodes
Search MassBank
2
MetfragR
3
ID level

Fig. 2

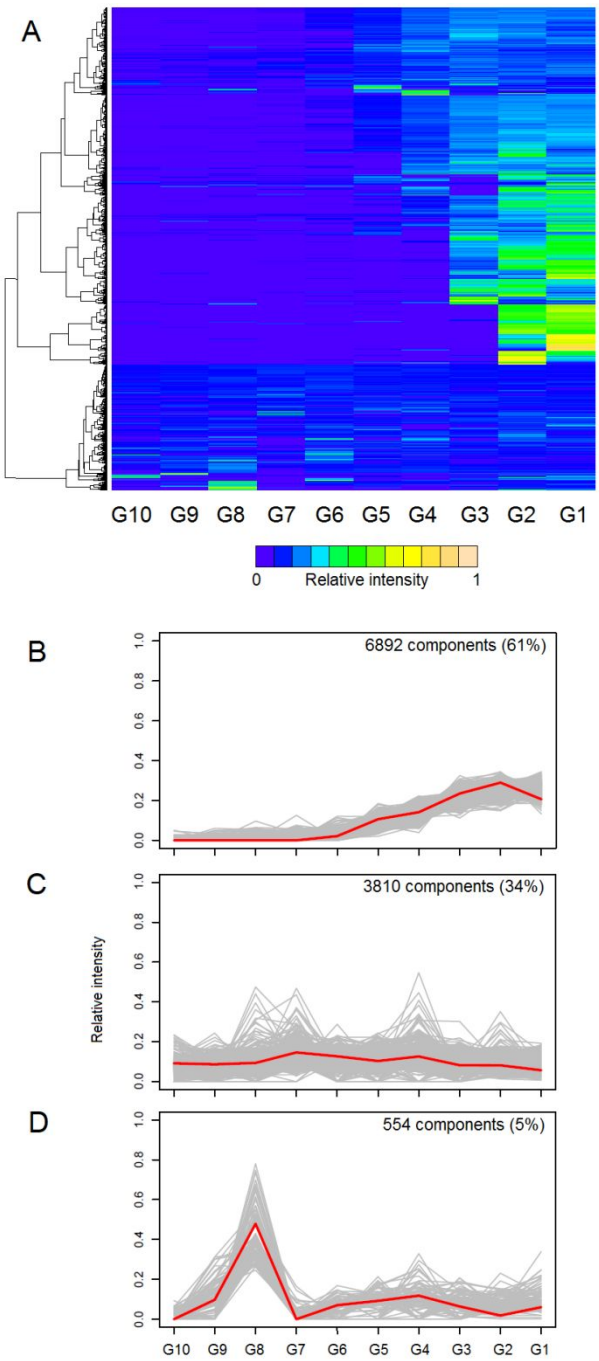

A
G10 G9 G8 G7 G6 G5 G4 G3 G2 G1
0
Relative intensity
1
B
6892 components (61%)
C
3810 components (34%)
Relative intensity
D
554 components (5%)
G10 G9 G8 G7 G6 G5 G4 G3 G2 G1

Fig. 3

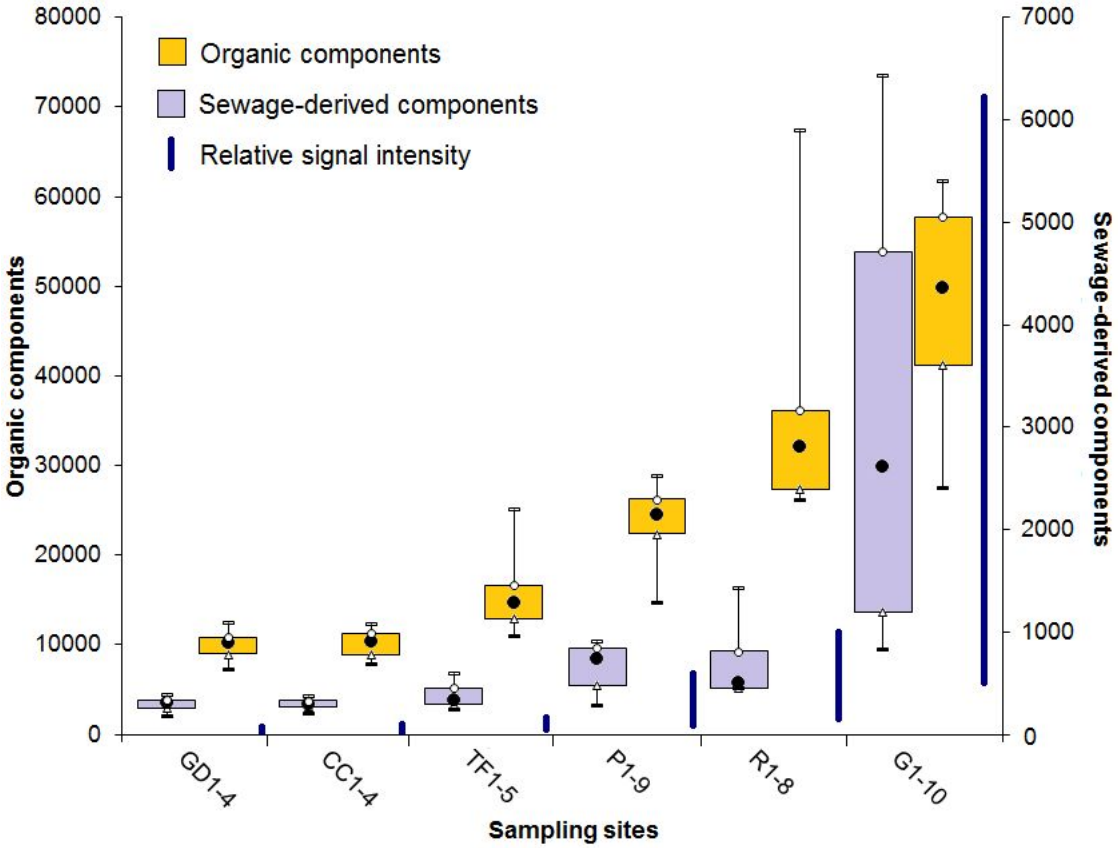

Organic components
Sewage-derived components
Relative signal intensity
Organic components
Sewage-derived components
80000
70000
60000
50000
40000
30000
20000
10000
0
7000
6000
5000
4000
3000
2000
1000
0
GD1-4
CC1-4
TF1-5
P1-9
R1-8
G1-10
Sampling sites

Fig. 4

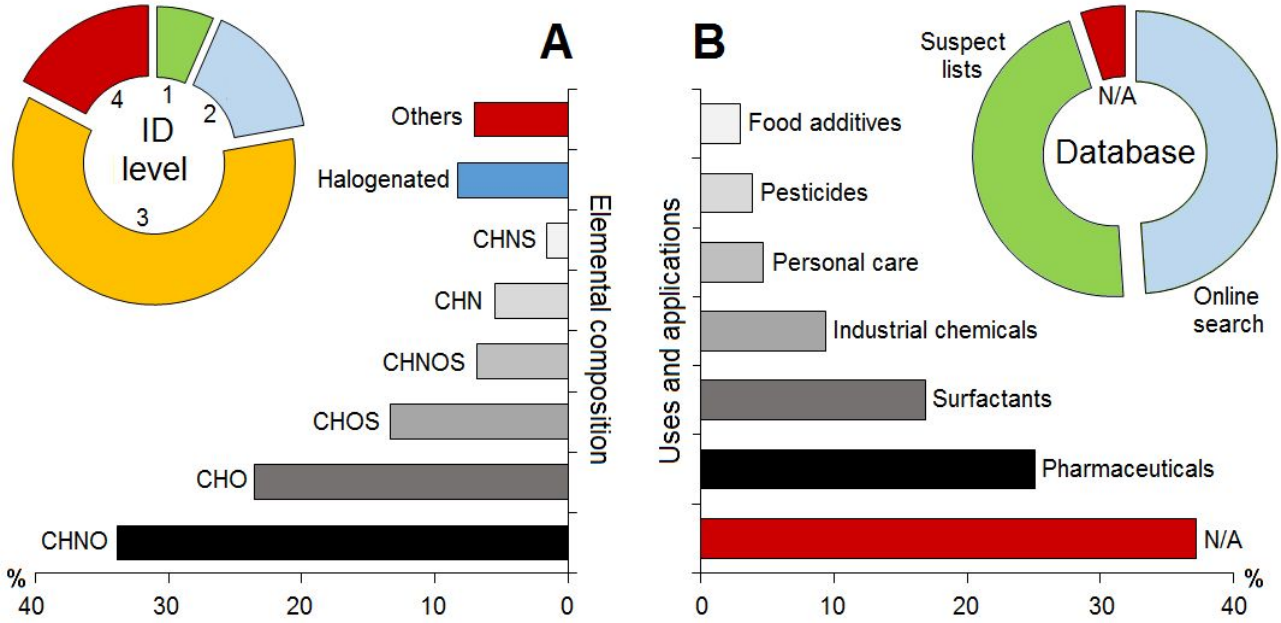

A
ID level
1
2
3
4
Others
Halogenated
CHNS
CHN
CHNOS
CHOS
CHO
CHNO
Elemental composition
%
40
30
20
10
0
B
Suspect lists
N/A
Database
Online search
Uses and applications
Food additives
Pesticides
Personal care
Industrial chemicals
Surfactants
Pharmaceuticals
N/A
0
10
20
30
40
%

Fig. 5

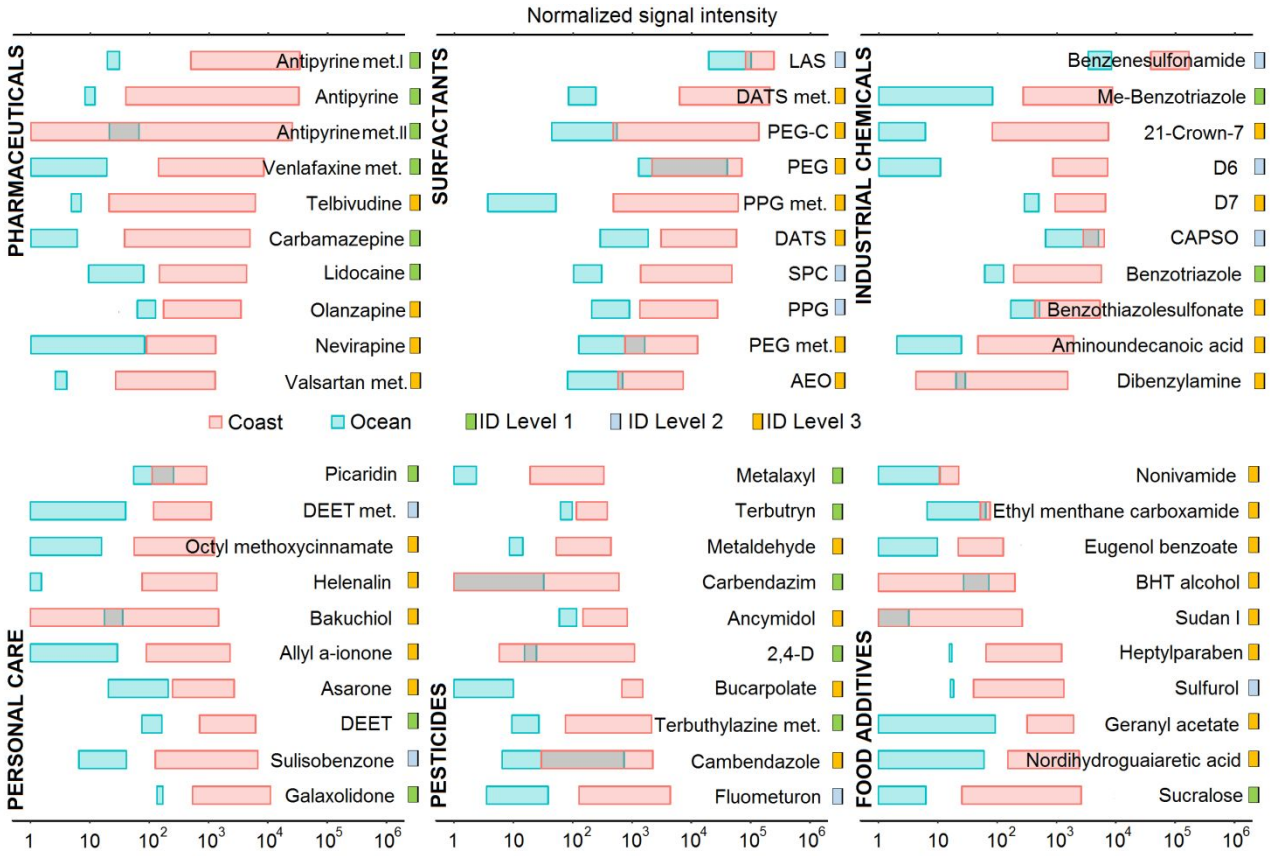

Normalized signal intensity
PHARMACEUTICALS
Antipyrine met.I
Antipyrine
Antipyrine met.II
Venlafaxine met.
Telbivudine
Carbamazepine
Lidocaine
Olanzapine
Nevirapine
Valsartan met.
SURFACTANTS
LAS
DATS met.
PEG-C
PEG
PPG met.
DATS
SPC
PPG
PEG met.
AEO
INDUSTRIAL CHEMICALS
Benzenesulfonamide
Me-Benzotriazole
21-Crown-7
D6
D7
CAPSO
Benzotriazole
Benzothiazolesulfonate
Aminoundecanoic acid
Dibenzylamine
Coast
Ocean
ID Level 1
ID Level 2
ID Level 3
PERSONAL CARE
Picaridin
DEET met.
Octyl methoxycinnamate
Helenalin
Bakuchiol
Allyl a-ionone
Asarone
DEET
Sulisobenzone
Galaxolidone
PESTICIDES
Metalaxyl
Terbutryn
Metaldehyde
Carbendazim
Ancymidol
2,4-D
Bucarpolate
Terbuthylazine met.
Cambendazole
Fluometuron
FOOD ADDITIVES
Nonivamide
Ethyl menthane carboxamide
Eugenol benzoate
BHT alcohol
Sudan I
Heptylparaben
Sulfurol
Geranyl acetate
Nordihydroguaiaretic acid
Sucralose
1
10
10^2
10^3
10^4
10^5
10^6

Fig. 6